Reclaiming the PCI Memory Range With Minimal Memory Loss in IA-32 Platforms

FIELD OF THE INVENTION

This invention relates generally to computer system programming, and more particularly to techniques for configuring computer system resources on IA-32 platforms.

BACKGROUND

Memory-Mapped IO. IA-32 processors permit applications to access input/output ("IO") ports in either of two ways: through a separate IO address space or through the physical memory address space. The latter approach is commonly referred to as "memory-mapped IO."

Address Space, Chipsets, and the BIOS. Beginning with the introduction of the Pentium Pro, IA-32 processors provide a 36-bit address bus. This enables them to support up to 64GB of physical memory. Accesses to this address space by the CPU and other devices within a computer are generally handled by several chips commonly known as the "chipset." A chipset provides bus interface, data path, instruction caching and similar functions on the motherboard. System firmware known as the basic input/output system ("BIOS") must configure the chipset at boot time with various information, including information about where main memory is located within the address space.

Chipset Configuration and the AGP Aperture. As a practical matter, the capabilities and requirements of the chipset dictate in large part where main memory can be located. The Intel E7505 chipset, for example, is designed to assume that certain devices will always be mapped into the peripheral component interconnect ("PCI") memory address range--that is, the address range beginning at 4GB and

extending downward far enough to include certain address regions. Specifically, it assumes that the advanced programmable interrupt controller ("APIC") addresses, the hub interface addresses, and any memory-mapped IO addresses will reside in this range. Another range of addresses the chipset assumes to exist in the PCI range is the accelerated graphics port ("AGP") aperture. By way of background, the operating system allocates numerous 4KB pages in main memory for use by the graphics controller. These allocated pages are normally discontiguous; but the graphics controller needs contiguous memory. So, a translation mechanism is used to create a series of contiguous logical addresses (the AGP aperture) within the PCI address range. These addresses map to the discontiguous pages that are allocated in main memory, wherever those might be.

By way of further background, configuration of the Intel E7505 and similar chipsets requires among other things that the BIOS write appropriate values into the top of lower memory ("TOLM") register and into the DRAM row boundary 7 ("DRB7") register. The TOLM register is designed to contain the maximum address below 4GB that should be treated as main memory. The DRB7 register is designed to contain the maximum address in the machine that should be treated as main memory. Thus, for machines having less physical memory than 4GB minus the minimum size required for the PCI memory address range, the TOLM and DRB7 registers will contain the same value. But for machines having more memory than that, the physical memory must be split because the PCI memory address range may not be moved. In such machines, there will be one region of physical memory located below the PCI range, and another region of physical memory located above the PCI range. The TOLM register will indicate the highest address within the first range. The DRB7 register will indicate the highest address within the second range.

Memory Reclaiming. In previous generation chipsets, any physical memory that was overlapped by the logical address space allocated to the PCI range was unusable. But in some workstations, the amount of memory allocated to memory-mapped IO devices could easily exceed 1GB. An unacceptably large amount of physical memory was rendered unusable in such machines.

Now, chipsets such as the E7505 attempt to provide a capability for "reclaiming" physical memory that is overlapped by the PCI range. In theory, this is done by remapping the physical memory lying within the PCI range to an equivalent-sized logical address range located just above the top of physical memory. During chipset configuration, the bottom and top of a remapped window are defined by values written by the BIOS into a REMAPBASE register and a REMAPLIMIT register, respectively. During normal operation, each incoming logical address is checked to determine whether it falls in the remapped window. If so, then the incoming logical address is remapped to the physical memory starting at the address defined by the TOLM register.

The AGP Aperture Bug. Unfortunately, the memory reclaiming feature of the current E7505 chipset operates erroneously: When a device writes to an address within the remapped window, the write is executed correctly. But when a device reads from an address within the remapped window, the E7505 chipset sometimes returns data from reclaimed physical memory, and other times returns data from a corresponding address within the AGP aperture! Moreover, no warning is given to the consumer of this memory when the error occurs.

The straightforward approach to addressing this problem would be to avoid using the memory reclaiming feature of the E7505 chipset altogether. This approach would ensure that no memory accesses could result in errors. The downside of the straightforward approach, however, is that it results in losing a large range of

addresses as unusable memory--a range equal in size to the entire PCI memory address range. As was mentioned above, this range can exceed 1GB in some workstations.

SUMMARY OF THE INVENTION

A preferred embodiment of the invention includes a method for configuring resources in IA-32 computers. For computers using the E7505 or a similar chipset, the inventive method enables the PCI memory address range to be reclaimed with minimal loss of available physical memory. In one aspect, the method includes: establishing a remap window at the top of physical memory, the remap window corresponding to the PCI memory address range; and reporting to the operating system that a portion of the remap window is reserved, the reserved portion corresponding to the AGP aperture within the PCI memory address range. In a further aspect, the method may include reporting to the operating system that the remainder of the remap window is usable memory.

On one hand, the inventive method configures the computer's resources dangerously: It enables devices in the computer to access the portion of the remap window that corresponds to the AGP aperture. If they should do so, they would encounter erroneous data when reading from that area of the address space. But on the other hand, the inventive method labels the remapped AGP aperture in such a way that it never actually will be accessed. (The area is marked as "reserved," but has no actual owner.) The result is that all of the PCI memory range except the AGP aperture is successfully reclaimed for general use.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a flow diagram illustrating a method for configuring IA-32 computer resources according to a preferred embodiment of the invention.

Fig. 2 is a block diagram illustrating the logical address space and physical memory as it would exist after executing the method of Fig. 1.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig. 1 illustrates a method 100 for configuring IA-32 computer resources according to a preferred embodiment of the invention. In one embodiment, the method was implemented in the BIOS. In alternative embodiments, the method may be implemented in other firmware or in software or hardware. Fig. 2 illustrates how the logical address space and physical memory will appear after the method of Fig. 1 has been executed. The inventive method will be described in relation to both Figs. 1 and 2.

In step 102, the BIOS determines the size of the PCI memory address range for the host computer. Normally, this will be the range of addresses between TOLM and 4GB. In addition, the BIOS should determine the starting address and size of the AGP aperture within the PCI memory address range. This may be accomplished by reading the values in the APBASE and APSIZE registers, respectively.

In step 104, the BIOS establishes a remap window at the top of physical memory, indicated in the drawing as TOM. This may be accomplished by writing the address of the top of physical memory (normally the value in the DRB7 register) into the REMAPBASE register, and writing this value plus the size of the PCI memory address range into the REMAPLIMIT register. Thus, the size of the remap window established will be equal to the size of the PCI memory address range.

The BIOS should also determine a translated AGP aperture address 200. Translated AGP aperture address 200 should be equal to APBASE - TOLM + REMAPBASE.

When the operating system boots, it normally queries the BIOS to determine which regions of the address space are valid for accesses. This may be done using the well-known interrupt 15h function e820h, or by alternative means. As indicated at step 106, the BIOS should report three memory ranges above 4GB in response to these operating system queries. Range 1 should extend from 4GB to translated AGP aperture address 200. Range 1 should be reported as usable memory. Range 2 should be equal in size to the AGP aperture (APSIZE), and should begin at translated AGP aperture address 200. Range 2 should be reported as "reserved" memory. Range 3 should begin at the top of range 2 and extend to REMAPLIMIT. Range 3 should be reported as usable memory.

After the memory above 4GB has been reported to the operating system in this manner, all of the PCI memory address range except the AGP aperture will have been reclaimed, as indicated in the drawing at 202. Any memory accesses directed to a logical address in the remap window will be remapped to a corresponding physical address within reclaimed area 202. The only unusable portion of physical memory remaining in the PCI memory address range will be region 204, which region will have the same size and location as the AGP aperture.

While the invention has been described in detail in relation to preferred embodiments thereof, the described embodiments have been presented by way of example and not by way of limitation. It will be understood by those skilled in the art that various changes may be made in the form and details of the described embodiments, resulting in equivalent embodiments that remain within the scope of the appended claims.